Title
Predicting Second and Third Graders' Reading Comprehension Gains: Observing Students' and Classmates Talk during Literacy Instruction using COLT.

Journal
Scientific studies of reading : the official journal of the Society for the Scientific Study of Reading, 24(5)

ISSN
1088-8438

Authors
Connor, Carol McDonald
Kelcey, Benjamin
Sparapani, Nicole
et al.

Publication Date
2020

DOI
10.1080/10888438.2019.1698583

Peer reviewed

Predicting Second and Third Graders' Reading Comprehension Gains: Observing Students' and Classmates Talk during Literacy Instruction using COLT

This paper introduces a new observation system that is designed to investigate students' and teachers' talk during literacy instruction, *Creating Opportunities to Learn from Text* (COLT). Using video-recorded observations of 2^{nd}-3^{rd} grade literacy instruction (N=51 classrooms, 337 students, 151 observations), we found that nine types of student talk ranged from using non-verbal gestures to generating new ideas. The more a student talked, the greater were his/her reading comprehension (RC) gains. Classmate talk also predicted RC outcomes (total effect size=0.27). We found that 11 types of teacher talk ranged from asking simple questions to encouraging students' thinking and reasoning. Teacher talk predicted student talk but did not predict students' RC gains directly. Findings highlight the importance of each student's discourse during literacy instruction, how classmates' talk contributes to the learning environments that each student experiences, and how this affects RC gains, with implications for improving the effectiveness of literacy instruction.

Key words: Reading comprehension; Reading instruction; English Language Arts; Classroom discourse; Teacher talk; Student talk

Predicting Second and Third Graders' Reading Comprehension Gains: Observing Students' and Classmates Talk during Literacy Instruction using COLT

This study introduces a classroom observation coding system, *Creating Opportunities to Learn from Text* (COLT), that offers a way to record and analyze teachers' and students' instructional discourse moves (i.e., teacher and student talk) during literacy instruction. Understanding the kinds of teacher and student talk that individually and collectively contribute to students' gains in reading comprehension (RC) is critical because in the elementary years, students are in the early stages of learning how to learn from texts. Literacy instruction contributes to students' ability to learn independently, to acquire knowledge, and to enjoy literature (Shanahan et al., 2010). The National Assessment of Educational Progress (NAEP, 2017) reveals that 60% of fourth graders lag behind standards for proficiency in RC. This is despite the considerable efforts researchers and educators have made to identify effective methods for teaching reading comprehension in the early elementary years and to provide teachers with information about these methods (e.g., National Reading Panel Report (NICHD, 2000); Shanahan et al., 2010).

Systematic classroom observations are a promising way to identify characteristics of teachers' instructional practices that are related to student literacy outcomes (Chinn et al., 2001; Mashburn et al., 2008; Pianta et al., 2016; Pianta & Hamre, 2009). However, students' talk is considered indirectly in many of these systems with the assumption that teacher instructional moves have a direct effect on what students learn (Kawalkar & Vijapurkar, 2013). COLT offers an alternative way of thinking about discourse with the hypothesis that understanding multiple students' talk during literacy instruction is likely to be more predictive of learning than focusing solely on teachers' practices and instructional discourse. With this in mind, we designed COLT to contribute to our understanding of aspects of teacher and student discourse during literacy lessons that are likely to influence students' achievement in RC. The hope is that the results might inform the development of more effective instruction and teacher professional development while providing insights into the complex world of classroom dynamics.

Theoretical Framework

This study is informed by the lattice model (Connor, 2016), which builds on seminal models of reading including the Simple View of Reading (Hoover & Gough, 1990), the Lexical Quality Hypothesis (Perfetti, 2007), and the Landscape model (Rapp et al., 2007). Central to the lattice model is the influence of the classroom learning environment on students' learning. Instruction is likely to be more effective if it considers students' constellations of skills, including text specific, linguistic, cognitive, and social skills and the associations among them (Connor, 2016). Additionally, it takes into account the reciprocal and bootstrapping effects of these processes and holds that students play an important role in shaping the classroom learning environment (Connor et al., 2009). The lattice model provides the rationale for focusing on the talk of multiple individual students in the classroom.

Studying complex aspects of teaching and learning is best done through systematic observations of naturally occurring early elementary literacy lessons (Pianta et al., 2016), noting that talking is the medium of teaching and is critical for relaying cultural and scientific knowledge from one generation to the next. Thus, the nature and extent of teacher and student discourse during instruction generally impacts students' learning (Curby et al., 2009; Dwyer et al., 2016; Kawalkar & Vijapurkar, 2013; Lawrence et al., 2015; Murphy et al., 2009).

Observations of Literacy Instruction

The COLT system builds on and extends studies that have identified teachers' practices that contribute to students' performance on measures of RC. Some have focused on the discourse environment (e.g., Gámez & Lesaux, 2012, 2015) whereas others have focused on broad or general characteristics of classroom climate that affect literacy learning. For example, Danielson's *Framework for Teaching* includes two domains divided into eight components (e.g., "teaching as managing classroom procedures" and "creating an environment of respect and rapport" (Kane & Staiger, 2012, pp. 20, MET, 2012). Other systems designed to study teachers' instructional practices, have reported a positive association between teachers' practices and students' language and reading achievement despite wide variation in content and purpose (e.g., Chiang, et al., 2017; Dwyer, et al., 2016; Gámez & Lesaux, 2015; Kelcey & Carlisle, 2013; Michener et al., 2018; Taylor et al., 2003). For example, the Measuring Effective Teaching project (MET, Kane et al., 2012) used three different systems, including the CLASS (Hamre et al., 2013), to study teachers' literacy instruction, all of which accounted for a small but significant amount of the variance in student achievement. The results of these studies guided our efforts to identify dimensions and features of instructional practices that would be critical in the development of COLT.

The missing component of these studies is students. Teachers' practices can help us understand teaching, but they do not inform us about students' contributions to learning (Fenstermacher & Richardson, 2005). Observations of teachers' practices without a measure of students' role in these learning opportunities are likely to be incomplete. We address this problem through the development of a protocol for observing multiple students' talk as they participate in literacy lessons. This is a critical component of the COLT system. It seemed likely that teacher talk and student talk together should significantly predict students' gains in RC. For example, Chiang et al. (2017) reported that teachers who encouraged students' oral language spent large amounts of instructional time *talking with students*, which is a promising practice. The COSTI (Smolkowski & Gunn, 2012) explicitly examined teacher-student interactions during kindergarten literacy instruction across four dimensions: explicit teacher demonstrations, student independent practice, student errors, and teacher corrective feedback. Student independent practice, which the authors described as “conceptually similar to opportunities to respond” (p.317) predicted their reading skills. Extending this research, we anticipated that data collected with the COLT system would provide insights into the role of student talk with their teachers and peers during literacy lessons. The authors of the *Individualized Classroom Assessment Scoring System* (inCLASS; Downer et al., 2010), which observes an individual student, noted that children’s interactions with teachers, peers, and tasks suggest the extent to which they are “building effective social relationships and acquiring skills/knowledge through instructional opportunities” (p. 1). Talk may be a major part of this process. Thus, we developed the COLT-Teacher and COLT-Student simultaneously with each informing the other.

The Components of COLT-Student

COLT-Student represents the major innovation of our endeavor. To gather insights into students’ involvement in early elementary literacy lessons, we built on theory and previous research to select promising dimensions and features of student talk in literacy lessons. One assumption was that active participation is central to students’ learning (Beck & McKeown, 2001). Ideally, teachers facilitate not just the mastery of basic reading skills but also ways of learning that depend on dialogic interactions with teachers and peers (Chiang et al., 2017; O'Connor & Michaels, 1993; Wells, 2007).

We identified three dimensions that reflected varying levels of talk from simple participation to generating ideas and questions, to interacting through discussion (Bloom et al., 1956; Chi, 2009; Connor et al., 2012; Hamre et al., 2013; Snow, 2010): *participating*, *generative*, and *interactive* talk (see Table 1 for detailed codes). For each dimension, we identified three or four specific types of talk that were supported both theoretically and empirically.

The first dimension, *participating talk*, focused on students' willingness to respond to their teachers' talk during lessons. Teachers often start lessons with questions that engage students' attention and interest. At the simplest level, student responses may indicate only that they are doing what they were asked to do. Three specific types of talk capture this basic level of participation: *non-verbal responding*, *answering simple questions*, and *reading text aloud*. Nonverbal responding might entail students' raising a hand if they liked the story they just read. Similarly, participating in group oral responses (e.g., reading a poem together) can also foster attention and interest. Quite commonly, teachers start a lesson by asking simple yes-no questions, using evaluation of students' responses (e.g., praise) to encourage participation (e.g., Dwyer et al, 2016). Although "known-answer" questions have often been criticized, student responses to such questions can initiate more cognitively engaging exchanges (e.g., Boyd & Rubin, 2006). O'Connor and Michaels (2007, p. 281) suggest that teacher questions that involve simple responses play a role in "socializing" children to patterns of interactions that draw them into their reading lessons. These types of talk were expected to be ubiquitous in classrooms.

The second dimension, *generative talk*, reflected higher-level talk in which students generate new ideas and make contributions to classroom discourse. Students' generative talk reflects their efforts to make sense of the topic or some part of the teacher's lesson—that is, forms of active engagement (Beck & McKeown, 2001; Chiang et al., 2017). We identified four types of talk that reflect this dimension (see Table 1). Three of these are central indicators of student involvement in group literacy discussions (e.g., Chinn et al., 2001; Clark et al., 2003; Murphy et al., 2009): *Answering questions that require thinking and reasoning*; *asking on-topic questions; using text to justify a response*. These reveal different facets of students' ability to analyze texts, generate arguments on topics, and draw inferences or conclusions. They reflect the Common Core Standards (2010). The fourth type of generative talk is *generative participation that does not follow classroom norms (i.e., off topic talk).* This type of talk is generative in that students are producing ideas, but the content is not directly related to the topic of the classroom discussion. We expected that generative student talk would be infrequently observed but, with the exception of off-topic talk, more likely to predict students' RC gains than participating talk because they reflect students' thinking (Chiang et al., 2017; Dwyer et al., 2016; Soter et al., 2008; Wolf, Crosson & Resnick 2005).

The third dimension, interactive talk, involved sustained interactive participation that might occur during group discussions, including exchanges among students (Almasi et al., 2001; Clark et al., 2003; Murphy et al., 2009). Two types of student talk reflected interactive talk: *participating in a discussion* and *voicing a disagreement*. The student had to be carefully attending to and participating in classroom discussion (Chinn et al., 2001). We hypothesized that interactive talk would be observed infrequently but would still predict students' RC gains.

Understanding the dynamics of the instructional discourse environment, including peer effects (e.g., Justice et al., 2011), is critical if we are to improve reading instruction for all students (Connor et al., 2009; Wolf et al., 2005). In addition to examining individual student talk, we examined whether the amount and type of multiple students' talk (i.e., classmate talk) in the classroom overall predicted students' RC gains.

The Components of COLT-Teacher

In developing the COLT-Teacher system, we built on the findings of studies that focus on teachers' discourse practices in elementary classrooms, particularly emphasizing those that predicted students' reading achievement in other observation systems. Across studies (Dwyer et al., 2016; Michener et al., 2018; Taylor et al., 2003; Wolf, Crosson & Resnick, 2005), promising aspects of teachers' talk include ways to engage students' interest and involvement in discussion, explanations and clarifications to support extended discussion, and requests to use thinking and analytic skills.

Comparison of study results can be challenging because researchers use somewhat different constructs to characterize aspects of teacher talk. For example, Michener et al. (2018) uses the term “teacher explanations” in a way that might be similar to Taylor et al.’s (2003) term “telling.”

Observation systems, such as COLT, must be sensitive to the context in which they are applied. In the early elementary years, teachers are introducing children to culturally valued reasoning practices. Not surprisingly, teacher practices include a fair amount of “scaffolding”-–that is, supporting students’ engagement in aspects of literacy learning that might seem to be just beyond their current reading and cognitive development. Thus, along with observing what teachers are teaching (i.e., reading skills and strategies), we saw a need to observe discourse practices that offer guidance and support for individual students and groups of students—emotional, social aspects of learning to read (Jadallah et al., 2011). With this in mind, we selected the following dimensions for the COLT-Teacher system: encouraging participation, facilitating extended talk, prompting students’ reasoning, and building knowledge (e.g., Chiang et al., 2017; Dwyer et al., 2016; Taylor et al., 2003; Wolf et al., 2005).

The first dimension, *encouraging participation*, represents commonly observed types of teacher talk for the purpose of engaging students’ attention and interest in the lesson and generating initial participation. These are discourse moves that draw students into the topic of the lesson. They might be very simple devices to stimulate attention, as might sometimes be necessary with young students. We identified two types of teacher talk for this dimension: *asking questions that require non-verbal student participation* (e.g., thumbs up if you agree) and *expressing interest in students' responses and ideas,* which should encourage students to keep participating (Michener et al., 2018).

The second dimension is *Facilitating extended talk*; it includes three types of teacher talk that should encourage participation and, perhaps, generative and interactive types of student talk. The three types are *inviting students to share information; summarizing or recasting student responses;* and *asking follow-up questions to gain information or clarify an idea.* Again, in some analyses of classroom discourse, questions that target student recitation and response to simple questions are viewed as having a negative influence of students' engagement and interest (Cazden, 2001), but subsequently several studies of classroom discourse have shown that recitation questions can actually trigger extended student discussion (Boyd & Rubin, 2006; O'Connor & Michaels, 1993; Wells, 2007). In short, teachers choose to use different types of questions to facilitate extended talk—for example, simple questions to introduce a topic but challenging "thinking" questions to clarify, extend, or deepen students' understanding of the topic. Boyd and Rubin (2006, p. 141) refer to the "strategic, targeted ways" that teachers use to engender student participation. Carlisle et al. (2013) found that teachers who frequently asked students to respond to short-answer questions about word meanings were also most likely to engage students in more extensive discussions (see also, Chiang et al., 2017).

The dimension *prompting students to reason* includes three types of teacher talk that should encourage students to use more generative types of talk. They include *challenging students to reason or draw conclusions about text; directing students to provide evidence from text; and engaging students in close analysis of text* (Michener et al., 2018). "Students need guidance in building and weighing arguments with warranted evidence, which requires that they clearly explicate their reasoning so that others can understand and build upon or critique their ideas." (O'Connor & Michaels, 2007, pp. 284-285). Research indicates that some infrequently observed or rare teacher discourse moves are associated with students' reading achievement (Carlisle et al., 2013). Taylor et al. (2003) found that even modest levels of higher-order questioning were associated with growth in students' reading. In addition, these types of teacher talk align with the Common Core Standards (2010) and a summary of recommendations for effective practices in teaching reading comprehension (Shanahan et al., 2010).

The dimension *building knowledge* directly relates to the research that shows that general and specific understanding about the topics they are reading support students' comprehension (Dwyer et al, 2016; Michener et al., 2018). This dimension includes three types of teacher talk that is specifically designed to build students understanding and knowledge:

explaining literacy concepts; encouraging students to make self-text or text-text connections, which should encourage generative talk, and *providing background information and content.* They are part of the Common Core standards (2010).

Summary and Research Questions

Our hypothesis is that understanding students' and teacher's talk will help us understand effective instructional discourse practices that support a strong classroom learning environment, which, in turn, promotes students' developing literacy skills, particularly RC. The following research questions guided our inquiry:

1. What is the nature of and variability in teacher talk observed during literacy instruction in second and third grade classrooms? To what extent are the types of teacher talk multidimensional? We hypothesized that teacher talk would be multidimensional following our four dimensions.
2. What is the nature of and variability in student talk observed during literacy instruction in second and third grade classrooms? To what extent are the types of student talk multidimensional? We anticipated that student talk would also be dimensional.
3. To what extent does teacher talk predict student RC outcomes?
4. To what extent do students' and classmates' talk predict RC outcomes? To what extent do students' and classmates' talk mediate the association between teacher talk and student outcomes?

Specifically, we hypothesized that teacher talk should influence the way or ways that individual students and their classmates talk during instruction. In turn, student and classmate talk should predict students' gains in RC over the course of one school year.

Methods

Our initial examination of the dimensions and codes was carried out through multiple reviews of observations conducted in 12 classrooms where six classrooms had high student RC achievement and six had weaker achievement. We created talk codes based on extant research coupled with multiple viewings of the 12 classrooms to create new codes, identifying teacher and student talk codes that were salient and appeared to differentiate classrooms. All of the student codes were created through this process. Iterative analyses of the observation data showed that some of the codes were highly correlated and captured similar types of talk; these were combined. Codes that did not predict RC outcomes were pruned unless they were of theoretical importance. The version of COLT presented here is the culmination of multiple iterations of development.

Participants

The classroom observations and student achievement data for this study were originally collected as part of a longitudinal study focused on literacy and math achievement (Authors). Please see *Appendix A* for the power analysis. To select our sample of students, we rank ordered students by their reading scores and then randomly selected six students – two from the lowest tercile, two from the middle, and two from the highest. Students in second (*n* = 175) and third grade (*n* = 162), and their teachers (*n* = 25 second; 26 third grade) across five schools in northern rural Florida from 2009–2011 participated in the current study. Approximately 25% of the students in the sample qualified for the National School Lunch Program (NSLP). Fifty-seven percent were girls, 83% identified as White, 6% African American, 4% Hispanic, 1% Asian, 1% Native American, and 5% multiracial. Approximately 10% of the students qualified for special education.

Participating second and third grade teachers were 96% female and reported an average of 15 years (*SD* = 9.21) of teaching experience. Ninety-seven percent identified as White and 3% identified as African American. All teachers had at least a B.A. or B.S. degree; 35% of the teachers reported having a Masters level degree.

Procedures

The classroom observations were conducted at the teachers' convenience during the district mandated 90-minute uninterrupted block of time devoted to literacy instruction. Each classroom's literacy block was video-recorded three times during the school year (fall, winter, and spring) with the exception of one second grade classroom, which was video-recorded only in the fall so 151 observations in all. Schools and teachers were consistent in providing reading instruction during the scheduled 90-minute block of time. Classroom video observations were coded in the laboratory using the Noldus Observer® Video-Pro Software (XT 11.5). We divided each of the classroom video observations of the literacy block into *lessons* (Dwyer et al., 2016). Lessons were our unit of analysis within classrooms and represent the planned divisions within the daily schedule of the literacy block—with each division having an overarching purpose and focusing on a specific learning activity. Once the 682 lessons were identified for the 151 observations in the 51 classrooms, six trained research assistants coded teacher and student talk using the COLT observation system. Interrater agreement was strong: 0.87 (Kappa) for the COLT-Teacher system and 0.90 (Kappa) for the COLT-Student system. Additional information is in *Appendix B*. A coded transcript is provided in Table C.1 in *Appendix C*. Coding manuals are available upon request from the corresponding author. During the literacy lessons, teachers used the district mandated curriculum, Houghton-Mifflin, and other materials, including both narrative and expository texts as well as leveled texts and teacher-prepared materials.

Standardized RC Measures

We used three measures of RC, which were administered in the fall and spring of the school year (see Table 3). These were the Passage Comprehension subtest of the *Woodcock-Johnson-III Tests of Achievement* (Woodcock et al., 2001) and the Reading Vocabulary and Reading Comprehension subtests of the *Gates MacGinitie Reading Test* (MacGinitie & MacGinitie, 2006). These are widely used standardized measures of reading that have consistently displayed high reliability in empirical research (alpha

> .80) and are highly correlated. With these assessments, we developed a latent construct of RC using a common factor model.

Analytic Strategies

Teacher and Student Measurement Models (see *Appendix A*).

We developed measurement models for teacher and student talk by comparing the relative and absolute fit of several factor structures within a multilevel framework that nested lessons and observations within teachers or students (e.g., Kelcey & Carlisle, 2013).

Confirmatory Factor Analysis *(see Appendix A)*. For student talk, factor structures considered the following latent representations of student talk: (a) a unidimensional structure; (b) a two-dimensional structure; (c) a bi-factor structure that integrated (a) and (b) such that every type of student talk was reflective of a general factor and, secondarily, a participating factor or generative/interactive factor; and (d) a three-dimensional structure as indicated in the original dimensions of student talk. We considered similar factor structures for teachers.

Mediation Models *(see Appendix A)*. We used multilevel structural equation modeling (SEM) to test our hypothesis that teacher talk operates through student talk to improve RC. To estimate the degree of mediation, we drew on the product of the relations between the teacher-student dimensions and student talk dimensions and achievement. Using the bi-factor representation, we first considered the total mediation effect for each student factor (*general, participating, and generative/interaction*). The total mediation effects describe how teacher talk acts upon each of the student talk dimensions—at the student- or classroom-level—in ways that facilitate improvements in RC. Thus, each total mediation effect captures the extent to which teacher talk improves students' RC by promoting individual student talk and/or by promoting classmate talk (i.e., the benefit of peer talk). Similarly, we can conceptually outline the contextual, compositional, or classroom mediation effect that captures the potential role of classmate talk in shaping individual student outcomes (e.g., Pituch & Stapleton, 2012)

Results

Overall, these second and third graders were reading at grade level expectations with a mean total score on the *Gates-MacGinitie Reading Test* (GMRT) falling at the 61st percentile in the spring (M = 61.64, SD = 25.52). However, there was substantial variability with percentile rank ranging from 0 to 99. Spring reading scores were lower for second graders (GMRT mean PR = 55.49) than for third graders (GMRT mean PR = 66.49). In general, fall and spring scores were stable with a correlation from fall to spring on the GMRT of .86 ($p < .001$). We observed consistent use of the Houghton-Mifflin curriculum in the classrooms.

Teacher Talk

Returning to the first research question: What is the nature of and variability in teacher talk observed during reading lessons in second and third grade classrooms? To what extent are the types of teacher talk multidimensional? Mean frequency and standard deviations of the different types of teacher talk, proportion of the variability on observed talk types across lessons due to persistent teacher differences (i.e., intraclass correlation coefficient), proportion of lessons where the talk type was observed, and factor loadings are provided in Table 4. With an average of four to five lessons per observation, teachers were observed to use the most frequent types of talk between 25 and 30 times during a single observation. Notably, these types of teacher talk ranged from questions that engaged students' participation to questions that challenged students' thinking. However, the frequent use of some types of talk was clustered within specific teachers for some types but not for others (Table 4). For example, approximately 65% of the observed variation in the frequency with which teachers used the *explaining literacy concepts type* was attributable to persistent differences among teachers whereas virtually all of the observed variation in the frequency of use with *expressing interest in students' responses or idea* was attributable to lesson to lesson and classroom differences.

Our assessment of the underlying factor structure regarding teacher talk suggested that the data were best described by a single factor (Table C.2 in *Appendix C*). The results suggested that the one- and two-factor models fit similarly; however, under the two-factor model the dimensions were correlated at about 0.99. All eleven types of talk, with the exception of *asking questions that require non-verbal responses*, loaded fairly highly on the single factor. Less frequently observed moves tended to be paired with higher factor loadings. For example, for *summarizing,* above average use of this move strongly differentiated teachers.

Student Talk

Our second question asked: What is the nature of and variability in student talk observed during reading lessons in second and third grade classrooms? To what extent are the types of student talk dimensional? We found that nine types of student talk were salient in these second and third grade classrooms (see Table 5). Again, these ranged in frequency from highly frequent, for example, *non-verbal responding*, which occurred, on average, for a single child about five times during each lesson and ranged from 0 to 40 times. *Using text to justify a response* was observed on average less than once per lesson and ranged from zero to five times per lesson. This variability was notable across lessons. Similarly, the extent to which their use was clustered in students and classes varied heavily by type of talk. *Verbally answering simple “wh” questions* was heavily clustered in students and classes—that is, there was relatively little variation within a student across lessons in the frequency with which a specific student used this type of talk. In contrast, even though *non-verbal responding* was commonly observed, there was a relatively high level of within student variability across lessons in the frequency with which a specific student used it—just eight percent was attributable to stable differences among students, while another 22 percent of the variation was attributable to differences among classes.

Overall, we found that student talk was multidimensional. Our assessment of the performance of the factor structures suggested that all four models (see Table C.3) were plausible from an absolute fit standpoint. From a relative standpoint, the bi-factor structure, with a *general talk* factor and secondary *participating* and *generative/interactive* factors best fit the data. The table also includes how each type of student talk relates to the *general talk* factor, the *participating talk* factor, and the *generative/interactive talk* factor (*generative* henceforth) in the bi-factor model.

Each type of student talk loaded reasonably well onto the general talk factor. However, the specific types of talk that best discriminated among levels of this factor were (1) *voicing a disagreement*, (2) *off-topic talk*, and (3) *using text to justify a response*. We found less strength in our measure's ability to differentiate students on the secondary factors of *participating talk* and *generative talk*. For *participating talk*, *non-verbal and verbal responding* reflected differences among students reasonably well but *reading text aloud* did not. For *generative talk*, the evidence was much more complex. Although *voicing a disagreement* was a stronger indicator of *general talk*, use of this type of talk had virtually no significant relation with the *generative talk* dimension. Furthermore, two other types of talk demonstrated strong negative relations with the *generative* factor. Whereas *off-topic talk* had a strong positive loading onto the *general talk* factor, it had a negative loading onto the *generative talk* factor. These results suggest that *off-topic talk* is positive in one sense because it suggests students are participating in learning opportunities but negative in another sense because such moves may undermine truly generative and interactive talk.

Teacher, Student, and Classmate Talk Predicting Reading Comprehension Gains

Our third research question asked: To what extent does teacher talk predict student RC outcomes? The total standardized association of teacher talk on student reading achievement gains from fall to spring (i.e., residualized change) was .11 (a small effect) and statistically significant.

Our fourth and final research question asked: To what extent does student and classmate talk predict RC outcomes? To what extent does individual student and classmate talk mediate the association between teacher talk and student outcomes? Classmate talk is the mean of all six observed students. The overall results are presented in Table 6 and Figure 1.

When we considered the mediating role of student talk (see Table 6 and Figure 1), results revealed a strong and significant sequence of explanatory relations connecting teacher talk, students' general talk (the *general talk* factor), and students' RC gains. Specifically, a standard deviation increase in teacher talk was associated with a .41 standard deviation increase in student *general talk*. In turn, a standard deviation increase in student *general talk* was associated with a .15 standard deviation increase in RC gains, with .05 attributable to individual students and .10 attributable to classmates. Once student and classmate talk were added to the model, teacher talk no longer directly predicted student RC outcomes. Thus, our hypothesis was supported. Teacher talk appeared to facilitate student and classmate talk (the *general talk* factor) and, in turn, individual student and classmate talk together predicted students' gains in RC.

When we considered the secondary student talk factors, *participating talk* and *generative talk*, we found broken chains of associations linking teacher talk, student talk, and reading achievement (see Figure 1 & Table 6). We found that increased use of teacher talk did not clearly foster more classmate *participating talk*, yet more classmate *participating talk* did predict stronger student RC gains. Individual student's *participating talk* did not significantly predict outcomes. In contrast, teacher talk predicted more individual *generative talk*, but these increases were not associated with students' RC gains (see Figure 1 & Table 6). In sum, student and classmate talk, both individual *general talk* and classmate *participating talk* significantly predicted gains in RC skills with a total effect of .27.

Discussion

In this study, we introduced the COLT observation system. To the best of our knowledge, the COLT system is the first to explicitly consider the impact of multiple individual students' talk on their developing literacy skills. We identified nine types of student talk. These include: *non-verbal responding, answering simple questions, reading text, answering questions that require thinking and reasoning, asking questions, using text to justify a response, off-topic talk, participating in a discussion,* and *voicing a* *disagreement*. Notably, these types of talk ranged from simpler types of talk, which were often choral, to higher-level kinds of talk that required students to generate ideas and express them. COLT includes 11 types of teacher talk, which ranged from *asking questions that require non-verbal responses* to *challenging students to reason or inference about text*. Once we considered students' talk in our models, teachers' talk no longer directly predicted students' outcomes. Rather, more teacher talk predicted students' outcomes indirectly by increasing student and classmate talk. The standardized overall effect of student and classmate talk on RC outcomes was .27 – which is greater than many other observation systems (e.g., MET Project, 2013); The standardized effect of teacher talk alone on RC was .11, which aligns with many other observations systems that focus solely on the teacher and the global classroom environment. Thus, by observing multiple students individually, we were able to explain a significant amount of variability in RC scores. This is important and suggests that students who share a classroom may still experience very different learning opportunities. At the same time, classrooms where more of the observed students were talking were generally associated with stronger students' RC gains than were classrooms where only one or two of the students were doing all the talking.

Our hypothesis regarding student talk as multidimensional was only somewhat supported since we had hypothesized three dimensions (participating, generative, and interactive talk). We found that a bi-factor model showed the best fit to our data, with one strong general factor (the *general talk* factor) and two weaker sub-factors (the *participating talk* and *generative talk* factors), where generative talk combined the generative and interactive dimensions. Again, *general talk*, which included all of the discourse moves, predicted students' RC gains. So too did the sub-factor classmate *participating talk* (i.e., at the classroom level). The general factor for student talk suggests that many aspects of students' talk contribute to their learning – not just higher order discourse and discussion, as has been suggested in other research (e.g., Chiang et al., 2017; Murphy et al., 2009). Thus, a first step in creating an effective literacy instruction environment might be to encourage as many students as possible to engage in instruction both verbally and non-verbally.

The finding that teacher talk was unidimensional provides support for researchers who argue that teachers will use a wide range of discourse moves to support their students' understanding of text (Johnson, 2017; Michener et al., 2018; Wei et al., 2018; Wells, 2007). We did find that the rarer discourse moves predicted greater variability in student talk than more frequently observed talk types. For example, *summarizing students' ideas,* which included *scaffolding a student's presentation of ideas* was rarely observed but loaded strongly on the teacher talk factor. Carlisle and colleagues (2013) found that teachers' rare moves, such as encouraging discussion, were associated with a high probability of employing more commonly observed moves. Particularly during early elementary literacy instruction, teachers are likely to, for example, recast an idea for group discussion only after one or more students have responded to simpler short-answer questions (see also Michener et al., 2018; Wei et al., 2018). Future research is needed to examine this sequence (and other sequences) of teacher and student discourse moves.

Teacher talk did not significantly predict classmate *participating talk.* Yet, classmate *participating talk,* which did predict students' RC gains, included some of the most frequently observed student discourse moves (i.e., *non-verbal responding, answering simple questions and reading text aloud*). A*sking questions that require non-verbal responses* was among the most frequently observed teacher talk types. The corresponding student talk, *non-verbal responding*, was also observed frequently. These kinds of responses are frequently choral responses in which most students respond at the same time, so more students participate at any given time. Results suggest that teachers might use specific moves to promote classroom-wide talk, such as asking students questions that require non-verbal responding or asking students to read aloud as a group to increase students' opportunities to talk and engage.

Teacher talk predicted student (but not classmate) *generative talk*. Yet, neither student nor classmate *generative talk* predicted students' RC gains. Student *generative talk* was a weak factor with negative loadings. More teacher talk was associated fewer students *asking on-topic questions* and less *off-topic talk*. One might argue that student off-topic talk disrupts classroom discourse by derailing constructive discussion, thus, this talk probably should be suppressed. However, it is of concern that more teacher talk was associated with students asking *fewer* questions. Students *asking questions* represents student thinking and reasoning, which, research indicates, should be encouraged (Beck & McKeown, 2001; Chiang et al., 2017; Duke & Pearson, 2009; Kintsch, 2005). *Asking questions* loaded positively and highly on the student *general talk* factor (see Table 5), which predicted RC gains.

A key contribution of this study is the use of measurement and multilevel models, which allowed us to examine the variability in teacher, student, and classmate talk during instruction, at the lesson, classroom, and individual student level. Collectively, our findings leave open the possibility that at least some of the lesson-to-lesson and class-to-class variation in what students say may be driven by how teachers engage specific students in a given learning opportunity rather than attributable to fixed individual students' differences in terms of their propensity for certain types of talk (i.e., some children are talkative). Student characteristics might play an important role, but there is much we still do not know about teacher and student talk that, for example, initiate and sustain discussions—even though our and others' findings suggest the importance of student talk (Murphy et al., 2009; Wilkinson et al., 2010). COLT can be used to investigate this directly. When we looked more closely at the frequency of student talk within and between lessons and classrooms, across all nine types of student talk, about 75% of the variation in the individual student talk types were attributable to lesson-to-lesson differences within classrooms. About 10% of the variation was due to persistent student differences and the remaining 15% was due to persistent differences between classrooms (see Table 5). Thus, about 15% of the variability in student talk is likely teacher-driven (but could also be due to e.g., class composition differences). Thus, more variation in student moves is attributable to persistent classroom differences (arguably due to the teacher) rather than persistent student differences.

There are limitations that should be considered when interpreting these results. First, this study looked across contexts (e.g., whole class, small group) and content (e.g., code-focused and meaning-focused literacy instruction). We only considered contexts where teachers were actively interacting with students, which was about 75% of the 90-minute literacy block. Students completing seat work and times spent in transition and other non-instructional activities (Day et al., 2015) were not coded. It is notable that, on average, teachers were spending over an hour of the classroom literacy block in instruction in which the teacher and students were interacting (versus seatwork, for example). This is greater than the amount of time recorded in earlier studies (e.g., Foorman et al., 2006) and highly encouraging. Second, this is a correlational study so no causal claims should be made. Although we assume directionality in our analyses (i.e., teacher talk predicting student talk and student talk predicting outcomes), reciprocal and interacting effects are very likely given the dynamic nature of the classroom learning environment and would be suggested by the lattice model (Connor, 2016). Future research is needed to examine the dynamic and bi-directional interplay among teacher talk, student talk, and student reading outcomes. Third, our video observations were conducted from 2009–2011 in rural schools in northern Florida, prior to full implementation of the new standards set in the Common Core (2010), which encourage many of the student moves we observed using COLT (e.g., *using text to justify a response*). Thus, teacher and student talk in today's classrooms may differ substantially. Furthermore, these findings may not generalize to other classroom settings in districts with different demographics, curriculum, and policies. It is possible that different teacher discourse moves might be more predictive in different settings or with students with different learning needs. More research in more classrooms and in earlier and later grades is needed. Finally, our sample was not sufficient to run more complex models that may have been more optimal.

Looking toward possible educational implications, we suggest that teachers and coaches might use the results of our study to consider possible ways to encourage more student talk. The nine types of student talk and the eleven types of teacher talk highlighted in COLT would be a useful resource. The results suggest value in attending to individual student's talk as an integral part of the general discourse environment during literacy instruction. Teachers' analysis of their own teaching (perhaps through video or audio recordings) might offer them opportunities to examine the classroom as a dynamic learning environment per the lattice model. Paying attention to student talk is critical because students who share the same classroom do not necessarily have the same background or opportunities to talk and learn, with implications for their RC achievement; the contribution of classmate talk to the discourse environment is worth noting. Practices such *as turn and talk* and *think, pair, share*, which encourage more students to talk are promising. Finally, although more research is needed, including experiments, our findings reveal that it is not enough for teachers to talk to students. Rather, they need to talk with students in ways that actually stimulates students' talk, even if some of this talk is at a simpler level. When students and their classmates talk, they are generally more actively involved in literacy learning opportunities, which appears to lead to stronger RC achievement for all students.

References

Almasi, J. F., O'Flahavan, J. F., & Arya, P. (2001). A comparative analysis of students and teacher development in more and less proficient discussions of literature. *Reading Research Quarterly, 36*, 96-120.

Beck, I. L., & McKeown, M. G. (2001). Inviting students into the pursuit of meaning. *Educational Psychology Review, 13*, 225-241.

Bloom, B., Englehart, M., Furst, E., Hill, W., Krathwohl, D. N. Y., & Toronto:. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: Longmans, Green.

Boyd, M., & Rubin, D. (2006). How Contingent Questioning Promotes Extended Student Talk: A Function of Display Questions. *Journal of Literacy Research, 38*(2), 141-169. doi:10.1207/s15548430jlr3802_2

Carlisle, J. F., Kelcey, B., & Berebitsky, D. (2013). Teachers' Support of Students' Vocabulary Learning During Literacy Instruction in High Poverty Elementary Schools. *American Educational Research Journal, 50*(6), 1360-1391. doi:10.3102/0002831213492844

Cazden, C. B. (2001). *Classroom Discourse: The Language of Teaching and Learning (2nd ed.)*. Portsmouth: Heinemann.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science, 1*, 73-105. doi:10.1111/j.1756-8765.2008.01005.x

Chiang, H., Walsh, E., Shanahan, T., Gentile, C., Maccarone, A., Waits, T., . . . Rikoon, S. N.-W., DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. (2017). *An Exploration of Instructional Practices that Foster Language Development and Comprehension: Evidence from Prekindergarten through Grade 3 in Title I Schools* Retrieved from Washington DC: https://files.eric.ed.gov/fulltext/ED575194.pdf

Chinn, C., Anderson, R. C., & Waggoner, M. (2001). Patterns of discourse during two kinds of literature discussion. *Reading Research Quarterly, 36*(378-411).

Clark, A., Anderson, R. C., Kuo, L.-J., Kim, I.-H., Archodidou, A., & Nguyen-Jahlel, K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review, 15*, 181-198.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

Connor, C. M. (2016). A lattice model of the development of reading comprehension. *Child Development Perspectives, 10*(4), 269-274. doi:10.1111/cdep.12200

Connor, C. M., Day, S. L., Phillips, B., Sparapani, N., Ingebrand, S. W., McLean, L., . . . Kaschak, M. P. (2016). Reciprocal Effects of Self-Regulation, Semantic Knowledge, and Reading Comprehension in Early

Elementary School. *Child Development, 87*(6), 1813-1824. doi:10.1111/cdev.12570

Connor, C. M., Morrison, F. J., Fishman, B., Ponitz, C. C., Glasney, S., Underwood, P., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85-99. doi:10.3102/0013189X09332373

Connor, C. M., Rice, D. C., Canto, A. I., Southerland, S. A., Underwood, P., Kaya, S., . . . Morrison, F. J. (2012). Child Characteristics by Science Instruction Interactions in Second and Third Grade and Their Relation to Students' Content-Area Knowledge, Vocabulary, and Reading Skill Gains. *The Elementary School Journal, 113*(1), 52-75. doi:10.1086/665815

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C., . doi: . (2009). Teacher–child interactions and child’s achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology,, 4*, 912–925. doi:10.1037/a0016647

Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade classroom environment. *Journal of School Psychology, 53*(5), 409-428. doi:http://dx.doi.org/10.1016/j.jsp.2015.07.004

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS):

Preliminary reliability and validity of a system for observing preschoolers’ competence in classroom interactions. *Early Childhood Research Quarterly, 25*(1), 1-16.

Duke, N. K., & Pearson, P. D. (2009). Effective Practices for Developing Reading Comprehension. *Journal of Education, 189*(1-2), 107-122. doi:10.1177/0022057409189001-208

Dwyer, J., Kelcey, B., Berebitsky, D., & Carlisle, J. F. (2016). A Study of Teachers’ Discourse Moves That Support Text-Based Discussions. *The Elementary School Journal, 117*(2), 285-309. doi:10.1086/688896

Fenstermacher, G. D., & Richardson, V. (2005). *On making determinations of quality in teachingg*. Retrieved from Ann Arbor, MI: http://www-personal.umich.edu/~gfenster/teaqual14ss.PDF

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*, 1-29.

Gámez, P. B., & Lesaux, N. K. (2012). The relation between exposure to sophisticated and complex language and early-adolescent English-only and language minority learners’ vocabulary. *Child Development, 83*(4), 1316-1331.

Gámez, P. B., & Lesaux, N. K. (2015). Early-adolescents’ reading comprehension and the stability of the middle school classroom-language environment. *Developmental Psychology, 51*(4), 447.

Hamre, B. K., Pianta, R. C., Jason, T. D., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Hamagami, A. (2013). Teaching through Interactions: Testing a Developmental Framework of Teacher Effectiveness in over 4,000 Classrooms. *The Elementary School Journal, 113*(4), 461-487. doi:10.1086/669616

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127 - 160.

Jadallah, M., Anderson, R. C., Nguyen-Jahiel, K., Miller, B. W., Kim, I.-H., Kuo, L.-J., . . . Wu, X. (2011). Influence of a Teacher's Scaffolding Moves During Child-Led Small-Group Discussions. *American Educational Research Journal, 48*(1), 194-230. doi:10.3102/0002831210371498

Johnson, E. M. (2017). Balancing comprehension and conversation: How elementary teachers manage multiple purposes for text discussions. *Teaching and Teacher Education, 66*, 325-337. doi:https://doi.org/10.1016/j.tate.2017.05.005

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills. *Child Development, 82*(6), 1768-1777. doi:10.1111/j.1467-8624.2011.01665.x

Kane, T., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from http://eric.ed.gov/?id=ED540962

Kane, T., Staiger, D. O., & McCaffrey, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from

Kawalkar, A., & Vijapurkar, J. (2013). Scaffolding Science Talk: The role of teachers' questions in the inquiry classroom. *International Journal of Science Education, 35*(12), 2004-2027. doi:10.1080/09500693.2011.604684

Kelcey, B., & Carlisle, J. F. (2013). Learning About Teachers' Literacy Instruction From Classroom Observations. *Reading Research Quarterly, 48*, 301-317.

Kelcey, B., Dong, N., & Spybrook, J. (2018). Sample size determination in multilevel mediation. *Prevention Science*, 1-28.

Kelcey, B., Hill, H., & McGinn, D. (2014). Approximate measurement invariance in cross-classified rater-mediated assessments. *Frontiers in Psychology, 5*(1469), 1*24.

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes, 39*(2-3), 125-128. doi:10.1080/0163853X.2005.9651676

Lawrence, J. F., Crosson, A., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word Generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal, 52*(4), 750-786. doi:10.3102/0002831215579485

MacGinitie, W. H., & MacGinitie, R. K. (2006). *Gates-MacGinitie Reading Tests* (4th ed.). Iowa City: Houghton Mifflin.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., . . . al., e. (2008). Measures of classroom quality in prekindergarten and children’s development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

MET Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from MET Project*'*s three-year study*. Retrieved from Seattle, WA:

Michener, C. J., Proctor, C. P., & Silverman, R. D. (2018). Features of instructional talk predictive of reading comprehension. *Reading and Writing, 31*, 725-756.

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*(3), 740-764. doi:10.1037/a0015576

NAEP. (2017, January 2010). National Assessment of Educational Progress: The nation's report card. Retrieved from http://nces.ed.gov/nationsreportcard/

NICHD. (2000). *National Institute of Child Health and Human Development, National Reading Panel report: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and* *its implications for reading instruction* (NIH Pub. 00-4769). Retrieved from Washington DC:

O'Connor, M. C., & Michaels, S. (1993). Aligning Academic Task and Participation Status through Revoicing: Analysis of a Classroom Discourse Strategy. *Anthropology & Education Quarterly, 24*, 318-335. doi:10.1525/aeq.1993.24.4.04x0063k

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. doi:10.1080/10888430701530730

Pianta, R. C., Downer, J. T., & Hamre, B. K. (2016). Quality in Early Education Classrooms: Definitions, Gaps, and Systems. *The Future of Children, 28*(2), 119-137.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher, 38*, 109-119. doi:10.3102/0013189x09332374

Pituch, K., & Stapleton, L., 41, . (2012). Distinguishing between crossand cluster-level mediation processes in the cluster randomized trial. *Sociological Methods & Research, 41*, 630–670.

Rapp, D. N., van den Broek, P., McMaster, K., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 389-312. doi:10.1080/10888430701530417

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. K. (2010). *Practice guide: Improving reading comprehension in kindergarten through 3rd grade*. (NCEE 2010-4038). Washington DC: Department of Education.

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student–Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly, 27*(2), 316-328. doi:https://doi.org/10.1016/j.ecresq.2011.09.004

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*, 450-452. doi:10.1126/science.1182597

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriquez, M. C. (2003). Reading growth in high-poverty classrooms: The influences of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal, 104*, 3-28.

Wei, L., Murphy, P. K., & Firetto, C. M. (2018). How Can Teachers Facilitate Productive Small-Group Talk?: An Integrated Taxonomy of Teacher Discourse Moves. *The Elementary School Journal, 118*(4), 578-609. doi:10.1086/697531

Wells, G. (2007). Semiotic mediation, dialogue and the construction of knowledge. *Human Development, 50*(244-274), 100-103.

Wilkinson, I. A. G., Soter, A., & Murphy, P. K. (2010). Developing a model of quality talk about literacy text. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 142-169). New York, NY: Guilford.

Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2005). Classroom talk for rigorous reading comprehension instruction. *Reading Psychology, 26*, 27-53.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson-III Tests of Achievement*. Itasca, IL: Riverside.

Table 1

Dimensions of Student Talk within Reading Lessons

Dimension	Student Talk	Description
Participatory Talk	Non-verbal responding	The student **responds** to the teacher's on-topic questions or **initiates** non-verbal communication. Examples include, raising hands, thumbs up/down, nodding or shaking head in response to a teacher's question.
	Verbally answering simple questions	The student verbally answers simple "wh", yes/no, or choice questions with brief, factual responses or labeling. Often this includes answering close-ended questions (e.g., "Where did she go to buy the milk? "Did she buy milk or eggs?)
	Reading text aloud (chorally or individually)	Two or more students read connected text together as a group.
Generative Talk	Answering questions that require thinking or reasoning	The student answers questions that require thinking, reasoning, analysis, or synthesis. Questions must require the student to think beyond and **provide new information** from what is presented in the text. Often the student will generate an answer in response to open-ended questions (e.g., "Why do you think they let the man in?" "How are these animals alike" "What are some examples?").
	Asking on-topic questions	The student asks on-topic questions to the teacher or a peer during the reading lesson ("Why is the boy sad?").
	Using text to justify a response	The student uses text to justify a response with or without support from the teacher. For example, the teacher may say "let's all look at page 10, what evidence can you find here to support that statement?" or the student may take initiative without being prompted to reference the text and say "I know she likes gorillas because it says here that she studied them for over 40 years!"
	Off topic talk: Generative participation	The student is involved in generative participation that violates the classroom cultural norms—the student's contributions are

	that does not follow classroom norms	disruptive or inappropriate to the discussion. This may include times when the student blurts out off-topic generative comments.
Interactive Talk	Participating in a discussion	The student is an active, contributing member within a discussion, a cohesive exchange of ideas centered on a given topic. Discussion includes three complete turns.
	Voicing a disagreement	The student voices a disagreement by making a comment or asking a question that challenges the initial words or statement.

Table 2

Dimensions of Teacher Talk within Reading Lessons

Dimension	Teacher Talk	Description
Encouraging Participation	Asking questions that require nonverbal participation	The teacher asks the student (s) on-topic questions that require non-verbal responses (both choral and individual), such as “Raise your hand if…” “Did you like the story? Give me thumbs up/down.” Questions might also require students to underline, highlight, and copy, vocabulary words or text.
	Expressing interest in students’ responses/ideas	The teacher values the student’s ideas and provides feedback or praise in a genuine manner, expressing sincere interest in the student’s contribution. For example, the teacher states, “I really like how you used our new vocabulary word, fathom in your sentence.
Facilitating Extended Talk	Facilitating sharing of ideas and information by calling on many students and by encouraging students to freely call out ideas (*Inviting students to share information)*	The teacher encourages students to share ideas and information by calling on many students or letting students freely call out to contribute to a conversation or discussion about the text (s). For example, the teacher might ask the students to express their thoughts about a particular topic or another student’s idea.
	Summarizing students’ ideas or synthesizing	The teacher synthesizes or summarizes one or more students’ ideas to support comprehension and/or recasts a student’s idea to stimulate further discussion on a given topic. For example, the

	responses and/or recasting to the group (*Summarizing or recasting student responses*)	teacher supports group discussion about the text by taking up a student's idea, summarizing it, and providing an opportunity for the group to add to it.
	Asking follow-up questions to gain information or clarify an idea	The teacher asks questions or requests further explanation about a student's response or idea for the purpose of gathering additional information or clarifying what the student meant.
Prompting Students to Reason	Challenging students to reason or draw conclusions about text *(Challenging students to reason or draw conclusions)*	The teacher asks questions that challenge students to reason or inference about the text. For example, the teacher might ask, "Why do you think he did that?" or "Are there other ways to explain the character's motivation. This also includes times when the teacher asks students to draw conclusions after reading the text, such as "What did you learn from this text?" What can you conclude about this book?" "What was the moral of the story?" "What was the author trying tell us?"
	Directing students to provide evidence from text	The teacher asks students to find rationale or use the text to support a response that they have given. For example, if a student has offered a general statement, such as, "Mr. Smith does not like animals." The teacher might say, "What was said in the text that gave you this idea?
	Engaging students in close analysis of text	The teacher directs the student's attention on a specific section of the text, encouraging the use of context cues to draw meaning from the text. This is likely to involve teacher-guided analysis and discussion of the text—sometimes a small part of the text such as the title or a phrase. This code also includes analysis of illustrations (e.g., the teacher is helping the student interpret lines in a drawing that the illustrator put in to show a character in motion) and text features (e.g., table of contents, paragraphs and headings, charts).
Building Knowledge	Explaining literary concepts	The teacher provides explanations or asks students to help explain such literacy concepts as understanding "main idea" or "supporting

		details." This may involve explicit instruction along with illustrations or applications to a given text.
	Asking students to share experiences to encourage text-to-self connections or make text-to-text connections *(Encouraging students to make text-self or text-text connections)*	The teacher encourages students to make personal connections with the text (text-to-self) by asking students to think about how their own life or knowledge relates to the text (e.g., This story is about a farm—have any of you been to or worked on a farm before?). The teacher might also ask students to compare or contrast two different texts (text-to-text). For example, the teacher asks whether the topics covered in two nonfiction books on sea creatures are similar.
	Providing background information/content	The teacher provides information to students to foster their understanding of a text or the conversation. This might include explaining the meaning of a vocabulary word that they will encounter in the text or providing the students with important background information about the characters in a story prior to reading the text.

Table 3

Reading Comprehension Assessments

Assessment	Construct assessed	Description	Reliability
Woodcock-Johnson Tests of Achievement-III Passage Comprehension	Reading Comprehension	A close task, students read increasingly difficult passages of all kinds (i.e., narrative, expository) and provide the word that is missing from the passage.	Alpha = 0.83
Gates-MacGinitie Reading Test, Comprehension	Reading comprehension	Students read increasingly difficult passages of all kinds and answer multiple four choice questions that require simple remembering to inferencing and higher order thinking.	K-R 20 reliability coefficient = 0.92–0.93
Gate-MacGinitie Reading Test, Reading Vocabulary	Reading and Vocabulary	Students read a word in a short sentence and choose the correct meaning of the word from four choices.	K-R 20 reliability coefficient = 0.92–0.93

Note. K-R 20 = Kuder-Richardson Formula 20

Table 4

Usage and Factor Loading Coefficients for the Teacher Talk Factor

Teacher Talk Type	Mean Frequency of Moves (SD)	% of Lessons where Teachers' Talk Type Were Used		Loadings	ICC[a]
		Observed once	Observed ≥ 2 times		
Encouraging Participation					
Asking questions that require non-verbal responses	1.83 (4.04)	1.9	42.2	0.02	0.23
Expressing interest in students' responses/ideas	0.34 (0.64)	2.8	25.3	0.89	0.01
Facilitating Extended Talk					
Facilitating sharing of ideas and information by calling on many students and by encouraging students to freely call out ideas	1.30 (1.96)	0	9.6	1.00	0.12
Summarizing students' ideas; scaffolding a student's presentation of ideas to encourage further discussion; synthesizing responses from different students and recasting to group	0.15 (0.51)	0	0.1	1.62	0.09
Asking follow-up questions to gain information or clarify an idea	1.36 (2.06)	2.7	46.9	1.24	0.13
Prompting Students to Reason					
Challenging students to reason or inference about text and challenging students to draw conclusions about text (c18)	1.16 (1.98)	0	10.2	1.58	0.06

Directing students to use text to support responses/answers	0.21 (0.63)	1.3	9.1	0.68	0.08
Engaging students in close analysis of text	0.14 (0.46)	1.2	9.1	1.24	0.00
Building Knowledge					
Explaining literacy concepts	0.16 (0.54)	1.2	9.6	0.70	0.65
Asking students to share experiences to encourage text-to-self connections and asking students to make text-to-text connections	0.18 (0.64)	0	0.7	0.60	0.00
Providing background info. with contextual or informative content	0.10 (0.45)	0.7	5.7	0.36	0.33

Note. See Table 2 for explanation of each code. Frequency computed by lessons across classrooms with lessons lasting, on average, 15:73 minutes (*SD* = 10:64).
[a] We report the ICCs (intraclass correlation coefficients) of teacher moves only as a heuristic indicator of the clustering or the proportion of variance attributable to persistent differences among teachers in their use of moves across lessons.

Table 5

Usage of Student Talk Types and Factor Loading Coefficients and Variance for the Bi-factor Structure of Student Talk

Student Talk Type		% of Lessons Where Students' Types of Talk (i.e., Moves) Were Used				Loadings for Student Talk Types		
	Mean Frequency of talk	*Present once*	*Present >2 times*	*Student ICC*[a]	*Class ICC*[b]	General	Participatory	Generative
Active								
Non-verbal responding (raising hand, thumbs up/down, shaking head yes/no)	4.99 (5.92)	0.12	0.63	0.08	0.22	1	1	--
Verbally answering simple "wh ", yes/no, and choice questions (single child)	1.90 (4.29)	0.19	0.37	0.43	0.48	0.938	0.836	--
Reading text aloud	0.59 (1.72)	0.11	0.11	0.00	0.29	0.762	0.201	--
Generative								
Answering questions that require thinking or reasoning	0.35 (.90)	0.12	0.08	0.17	0.13	1.543	--	1
Asking simple, on-topic questions	0.09 (0.40)	0.05	0.02	0.00	0.02	1.327	--	-1.454

Using text to justify a response	0.03 (0.25)	0.01	0.01	0.00	0.19	1.814	--	1.599
Off-topic generative participation	0.07 (0.36)	0.04	0.01	0.05	0.00	1.954	--	-1.68
Interactive								
Participating in a discussion	0.08 (0.35)	0.05	0.01	0.18	0.09	1.55	--	0.135
Voicing a disagreement	0.00 (0.07)	0.01	0.00	0.00	0.00	2.597	--	0.03
TOTAL Mean Frequency Score (unscaled)	8.55 (0.22)							
Factor variance		--	--	--		0.26	0.29	0.10

Note. Frequency computed per lesson. Lessons lasted, on average for students 15:73 minutes (*SD* = 10:64). Resulting factor variances describe the persistent differences among students in their uses of moves across lessons.

[a] We report the ICCs (intraclass correlation coefficients) of student moves only as an indicator of the clustering or the proportion of variance attributable to persistent differences among students across lessons in the frequency of use for a type of talk.

[b] We report the ICCs (intraclass correlation coefficients) of student moves only as an indicator of the clustering or the proportion of variance attributable to persistent differences among classes across lessons in the frequency of use for a type of talk.

Table 6

Teacher Talk Predicting Student and Classmate Talk, Predicting Student Reading Comprehension.

Response	Predictor	Estimate	Std. Error
Student general talk	Student fall reading comprehension	0.10	0.07
	Classroom fall reading comprehension	0.15	0.21
	Teacher talk	0.41*	0.17
Student participating talk	Student fall reading comprehension	0.02	0.06
	Classroom fall reading comprehension	0.16	0.22
	Teacher talk	0.05	0.15
Student generative talk	Student fall reading comprehension	0.09	0.07
	Classroom fall reading comprehension	0.25	0.19
	Teacher talk	0.33*	0.13
Student spring reading comprehension	Student fall reading comprehension	1.00*	0.02
	Student general talk	0.05*	0.02
	Student participating talk	-0.03	0.03
	Student generative talk	0.01	0.02

Total fall reading comprehension	0.92*	0.06
Total general talk	0.15*	0.06
Total participating talk	0.12*	0.05
Total generative talk	0.06	0.06
Teacher talk	0.02	0.06

Note. Standardized path coefficients for the multilevel mediation model using student talk variables from the bi-factor model. See also Figure 1. Analyses treat the use of moves as ordinal in nature. Student codes: 0 not present, 1 present once, 2 present 2 or more times.
* $p < 0.05$

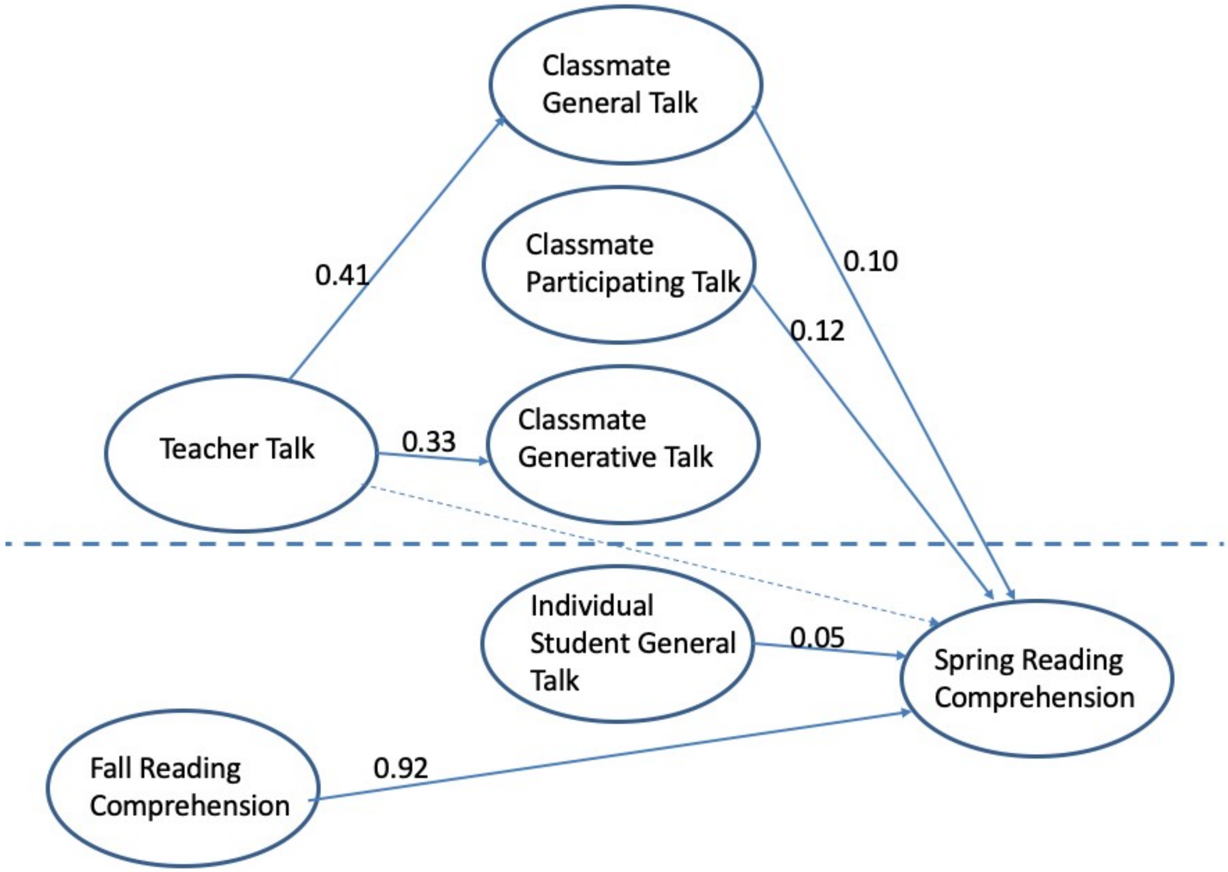

Figure 1. Path diagram (standardized path coefficients) of multilevel model of the effect of teacher and student/classmate talk on gains in reading comprehension using bi-factor model results (see Table 6). Solid lined arrows indicate paths significantly greater than 0 ($p < 0.05$) whereas dashed lined arrows are not significantly greater than 0. All variables are latent variables. Teacher talk includes 11 teacher moves; Student/Classmate talk includes 9 student moves; fall and spring reading comprehension (RC) each include scores from 3 assessments. Variables above the dashed line are classroom level variables whereas

variables below the dashed line are student level variables. Individual student participating and generative talk did not predict RC outcomes and so were not included in this figure. .

Appendix A: Analytic Strategies

Power Analysis

Practical implementation of classroom observation systems introduces construct-irrelevant variation or measurement error. For instance, past analysis of other systems has indicated that in addition to variation stemming from teachers and students, variation frequently arises from differences among, for example, raters, lessons, indicators, classroom settings, and their interactions. Synthesis of results in the literature suggest that teachers are likely to account for only about 1/3 of the total variance observed in instruction whereas the remaining 2/3 owes to construct-irrelevant sources such as variation among lessons, observations, and raters (e.g., Carlisle et al., 2013). Power analyses and design considerations suggested that with 50 teachers and six students per teacher, we would likely need at least three observations of the entire literacy block in order to maintain a level of reliability that would buttress a reasonable level of power.

Item Factor Analysis

For students, factor structures considered the following latent representations of student talk: (a) a unidimensional structure such that students' use of each type of talk is reflective of only a single factor; (b) a two-dimensional structure such that the types of observed talk were governed by a participating factor or a factor combined generative/interactive factor based on the original domains; (c) a bi-factor structure that integrated (a) and (b) such that every type of student talk was reflective of a general factor and, secondarily, a participating factor or generative/interactive factor; and (d) a three-dimensional structure such that use of each type of student talk was governed by only an participating factor, a generative factor, or an interactive factor as indicated in the original domains of student talk.

We considered similar factor structures for teachers. Specifically, we examined: (a) a unidimensional structure such that all teacher moves reflected a single factor; (b) a two-dimensional structure such that the types of talk was driven by an encouraging participation factor and prompting students to reason factor; and (c) a three-dimensional structure that further considered a facilitating extended talk factor (see Table 1).

To assess the relative and absolute fit of these structures, we used multilevel graded response model formulations. Using a bi-factor representation, the general form of our model was (e.g., Kelcey, Carlisle, & Berebitsky, 2013; Kelcey, Hill, & McGinn, 2014)

$$P(M_{ils}=k)=P(M_{ils}\geq k)-P(M_{ils}\geq k+1)$$
$$P(M_{ils}=k)=\frac{1}{1+\exp(-[a_i^G G_s+a_i^S S_s-d_i^{k-1}])}-\frac{1}{1+\exp(-[a_i^G G_s+a_i^S S_s-d_i^{k}])} \qquad (1)$$

Here M_{ils} is the use of type of talk *i* for student (or teacher) *s* in lesson *l*, a_i^G represents the general loading parameter for move *i* onto the general student (teacher) talk dimension (G_s; all types of talk) and a_i^S is the loading parameter for that same move onto its corresponding secondary dimension (i.e., S_s is *participating* [A_s] or *generative/interactive* [GI_s]). Let *K* represent the number of categories moves are graded on (three) with *k* as a specific category and let $d_i^1, \ldots, d_i^{K-1}$ be a set of *K*-1 ordered move thresholds. To identify the scale, the loading of the first move for each dimension was set to one.

Mediation Models. Our models mapped out the extent that teacher talk promoted student talk, which promoted students' reading comprehension. We assessed these associations using the following multilevel mediation model where *A* is student *participating talk*, *G* is student *generative talk*, and *I* is student *interactive talk*.

$$
\begin{aligned}
G_{sc} &= \pi_{0c}^{G} + \pi_1^{G} X_{sc}^{w} + \varepsilon_{sc}^{G} \quad \varepsilon_{sc}^{G} \sim N(0, \sigma_{G|}^2) \\
A_{sc} &= \pi_{0c}^{A} + \pi_1^{A} X_{sc}^{w} + \varepsilon_{sc}^{A} \quad \varepsilon_{sc}^{A} \sim N(0, \sigma_{G|}^2) \\
GI_{sc} &= \pi_{0c}^{GI} + \pi_1^{GI} X_{sc}^{w} + \varepsilon_{sc}^{GI} \quad \varepsilon_{sc}^{GI} \sim N(0, \sigma_{GI|}^2) \\
\pi_{0c}^{G} &= \zeta_{00}^{G} + a^{G} T_c + \zeta_1^{G} X_c^{b} + u_{0c}^{G} \\
\pi_{0c}^{A} &= \zeta_{00}^{A} + a^{A} T_c + \zeta_1^{A} X_c^{b} + u_{0c}^{A} \quad u_{0j} \sim MVN(0, \boldsymbol{\tau}),\ \boldsymbol{\tau} = \begin{pmatrix} \tau_{G|}^2 & & \\ \tau_{GA|} & \tau_{A|}^2 & \\ \tau_{GGI|} & \tau_{AGI|} & \tau_{GI|}^2 \end{pmatrix} \\
\pi_{0c}^{GI} &= \zeta_{00}^{GI} + a^{GI} T_c + \zeta_1^{GI} X_c^{b} + u_{0c}^{GI}
\end{aligned} \tag{2}
$$

The first three equations in (2) capture differences among students within classrooms (student-level) whereas the last three equations capture differences among classrooms. At the student-level, we used G_{sc}, A_{sc}, and GI_{sc} as the latent dimensions of student talk under the bi-factor representation, π_{0c} as the classroom-level intercepts for each of the student talk dimensions, $\pi_1^G, \pi_1^A,$ and π_1^{GI} as the coefficients capturing these dimensions' associations with the student-level component of prior reading comprehension X_{sc}^{w}, and ε_{sc} as the student-specific residual errors. At the classroom level, we used T_c as the single latent dimension for teacher talk for teacher j with a^G, a^A, and a^{GI} as the path coefficients capturing the relations between the teacher talk dimension and each of the student talk dimensions, X_c^b as classroom-level measure of prior reading comprehension with ζ_1 as the path coefficient, and u_{0c} as the normally distributed classroom random effects.

The mediator model was linked with a multilevel structural model for reading comprehension such that

$$
\begin{aligned}
Y_{sc} &= \beta_{0c} + b_1^G G_{sc}^w + b_1^A A_{sc}^w + b_1^{GI} GI_{sc}^w + \beta_1 X_{sc}^w + \varepsilon_{sc}^Y \quad \varepsilon_{sc}^Y \sim N(0, \sigma_{Y_1}^2) \\
\beta_{0c} &= \gamma_{00} + B^G G_c^b + B^A A_c^b + B^{GI} GI_c^b + c' T_c + \gamma_1 X_c^b + u_c^Y \quad u_c^Y \sim N(0, \tau_{Y_1}^2)
\end{aligned} \tag{3}
$$

In the outcome model, we use Y_{sc} as the outcome for student s in classroom c, β_{0c} as the classroom-specific intercept, G_{sc}^w, A_{sc}^w, and GI_{sc}^w as the within/student components of each student talk dimension (i.e., capturing variation among students within classrooms) with b_1 as the path coefficients describing the student-level conditional relationship between each student move dimension and the outcome, X_{sc}^w as a student-level component of the prior reading comprehension measure with path coefficient β_1, ε_{sc}^Y as the individual-level outcome error with variability of $\sigma^2_{Y_|}$. At the classroom level, we use γ_{00} as the overall intercept, *B* as path coefficients capturing the conditional total (classroom- and student-level) association between each student talk dimension and the outcome, *c'* as the path coefficient for the residual association between teacher talk and reading achievement, γ_1 as the path coefficient linking the cluster-level pretest (X^b_c) and the outcome, and u^Y_c as the normally distributed classroom-specific random effect with variance $\tau^2_{Y_|}$.

To estimate the degree of mediation, we drew on the product of the relations between the teacher-student dimensions (*a* coefficients in (2)) and student talk dimensions and achievement (*B* coefficients in (3)). Using the bi-factor representation, we first considered the total mediation effect for each student factor (*general, participating, and generative/interaction*). We define this as

Total Student General Talk (TGT): $TGT = a^G B^G$

Total Participating Talk (TAT): $TAT = a^A B^A$

Total Generative/interactive Talk (TGIT): $TGIT = a^{GI} B^{GI}$

The total mediation effects describe how teacher talk acts upon each of the student talk dimensions—at the student- or classroom-level—in ways that facilitate improvements in reading comprehension. Put differently, each total mediation effect captures the extent to which teacher talk improves students' reading comprehension by promoting individual student talk and/or by promoting classroom-wide talk (i.e., the benefit of peer talk).

Under the assumption that the effects of talk accrue similarly for students in classrooms, we can descriptively decompose the total mediation effects for each student dimension into components that specifically delineate the flow of stronger teacher instruction through improved student talk and through stronger classroom talk. We recognize that such decompositions is controversial in the literature because technically multilevel mediation describes covariances at the cluster level (e.g., Pituch & Stapleton, 2012). However, our decomposition is simply meant to provide additional descriptions of associations among variables across levels. First, we consider the unique student-level mediation effect that examines the degree to which the effects of increases in teacher talk on student achievement are transmitted through increases in individual student talk in each dimension (i.e., *general talk*). We can obtain estimates of the student-level mediation effects (Kelcey et al., 2018):

Student-level General Talk (TGT): $TGT = a^{G} b_{1}^{G}$

Student-level Participating Talk (TAT): $TAT = a^{A}\, b_{1}^{A}$

Student-level Generative/Interactive Talk (TGIT): $TGIT = a^{GI}\, b_{1}^{GI}$

Each of these effects quantifies the improvement in achievement that accrues as a result of changes in individual student talk produced by teacher talk when holding constant classmates' talk.

Similarly, we can outline the contextual, compositional, or classroom mediation effect that captures the potential role of peer (i.e., classmate) talk in shaping individual student outcomes. The unique classroom mediation effect focuses on the association of the latent classroom levels of student talk in each dimension and the latent levels of student comprehension beyond that which is supplied by the correlation between comprehension and differences in individual participation (as captured by b_1). In other words, the classroom mediation effects estimate the increment in student reading comprehension that accumulates as a result of changes in classmates' talk in a dimension produced by teacher talk when holding constant individual student talk in a dimension. Estimates of the unique classroom mediation effects are obtained as

Classroom Student General Talk (CGT): $CGT = a^G (B^G - b_1^G)$

Classroom Participating Talk (CAT): $CAT = a^A (B^A - b_1^A)$

Classroom Generative/Interactive Talk (CGIT): $CGIT = a^{GI} (B^{GI} - b_1^{GI})$

Appendix B: Observation Coding Protocol

Observation Procedures

Each classroom's 90-minute literacy block (*N* = 51 classrooms) was video-recorded three times during the school year (fall, winter, and spring) with the exception of one second grade classroom, which was video-recorded only in the fall so 151 observations in all. Two video cameras were used to record all activities taking place during the entire 90 minutes devoted to English language arts and reading instruction. One camera was focused on the entire classroom so that we could see the students because individual students were coded, and the other focused on the teacher. If the teacher was working with a small group of students, the camera was moved to focus on the small group. During the observation, the research assistant took careful field notes as well as recorded information so that each student could be identified when coding the video tapes. We usually recorded what each student was wearing and other distinguishing features. The research assistant also recorded when students left the classroom, activities that might not be clear on the video tape, and information about instructional materials being used.

Defining and Observing Lessons

We divided each of the classroom video observations of the literacy block into *lessons* (Dwyer et al., 2016). Lessons represent the planned divisions within the daily schedule of the literacy block—with each division having an overarching purpose (e.g., providing information about a literacy concept) and focusing on a specific learning activity. The beginning and end were typically bordered by transitions, such as teacher requesting that students close their books and shift their attention to the teacher. Trained research assistants segmented each classroom video observations into lessons and recorded the teachers' instructional focus, yielding an average Kappa coefficient score of .82. On average, lessons lasted 15:73 minutes (min:sec, *SD* = 10:64) and there were four to five lessons per video observation of the literacy block; 682 lessons across the 151 observations were identified. By nesting students within lessons, within classrooms, for our analyses, we could begin to examine how student talk varied by lesson (within and between lesson variance) as well as within and between classrooms.

Coding Procedures

ISI Coding. Classroom video observations were coded in the laboratory. Trained research assistants first coded each video observation using the *Individualized Student Instruction* (ISI) coding system (Connor et al., 2009) and Noldus Observer® Video-Pro Software (XT 11.5), which provided information on the amount of time spent and type of instructional activities students received (e.g., decoding, vocabulary, sustained silent reading, writing) and the context (e.g., whole class with the teacher, seat work) for each individual student. The ISI coding manual is available upon request from the corresponding author. From these codes, we identified all of the instruction in which the teacher was actively interacting with students. If the activity involved students working independently or only with peers, we did not include those lessons in the corpus of learning activities to be coded.

Defining Lessons. We then divided each observation into *lessons* or *instructional activities*. The lessons represent the planned divisions within the daily schedule—each division having an overarching purpose (i.e., providing information about a literacy concept) and focusing on a specific activity (e.g. Lessons had a beginning, middle and end. The beginning and end were typically bordered by transitions, e.g., students, close your books and put your eyes on me). Trained research assistants segmented the classroom observation into lessons and recorded the instructional focus, yielding an average Kappa coefficient score of 0.82. On average, lessons lasted 15.73 minutes (*SD* = 10.64) with between 4 and 5 lessons per observation. Thus, 682 lessons across the 151 observation were identified. By nesting students within lessons for our analyses, we could begin to examine how child talk varied by lesson (within and between lesson variance).

Coding with the COLT system. Once the 682 lessons were identified for the 151 observations in the 51 classrooms, six trained research assistants coded teacher and student talk using the COLT observation system, which was entered into the Noldus Observer® Video-Pro Software (XT 11.5). A coded transcript is provided in Table C.1 in *Appendix C*. Coding manuals are available upon request from the corresponding author. COLT-Teacher was used to code teachers' discourse moves (teacher talk) and COLT-Student was used to code students' discourse moves (student talk). Teachers were coded first and then 6 randomly selected students per classroom were coded. Interrater agreement for the COLT-Teacher and COLT-Student systems were calculated separately at the classroom level. Interrater agreement was 0.87 (Kappa) for the COLT-

Teacher system and 0.90 (Kappa) for the COLT-Student system. All coders had to be reliable with each other and to a gold standard coder. Interrater agreement among coders was also calculated using percent agreement midway through the coding and yielded an agreement score (Kappa) of 0.78 for the COLT-Teacher and 0.84 for the COLT-Student systems.

Appendix C

Table C.1
Example Transcript

Transcript	**Teacher**	**Students (A, B, & C)**
Teacher: The genre of this story is something that we haven't talked about before. We've talked about fiction, we've talked about non-fiction, and we've talked about realistic fiction. This story is a parody of a classic folktale. A parody is like a ridiculous imitation. What is this one making fun of?	Explaining literacy concepts	
Student A, B, and C are listening to the teacher		
The teacher encourages the students to freely call out their responses.	Inviting students to share information	
Student A: Looking at the wolf?		Verbally answering simple questions
Student B: Looking to see if they got caught?		Verbally answering simple questions
Student C: Shaking her head yes.		Non-verbal responding
Students A, B, and C were listening to the interaction when they were not responding.		
Teacher: What true story or real folktale is this one making fun of?	Asking follow-up questions	

Transcript	**Teacher**	**Students (A, B, & C)**
Student B: The 3 Little Pigs.		Verbally answering simple questions
Teacher: Right, the 3 little Pigs. But this one is a parody – an imitation. It is just funny and tells the story from a different point of view. So it's called a parody.	Explaining literacy concepts	
Student A: So this is funny?		Asking simple, on-topic questions
Teacher: Yeah, I think it's funny.		
Student B: Wait, so is it really true – what the wolf says?		Asking simple, on-topic questions
Teacher: You will have to read it to find out! What is the overall theme of the original 3 Little Pigs?	Challenging students to reason or draw conclusions	
Student A: The wolf keeps blowing and blowing and blowing down the pig's house.		Answering questions that require thinking or reasoning
Teacher: And the wolf is the bad guy! And he keeps coming after the pigs and they are all so scared. Then finally he comes to the brick house	Providing background information with facts or informative content related to the text	
Teacher: Can the wolf get into the brick house?		

Transcript	Teacher	Students (A, B, & C)
Students A, B, and C answering together: No!		Verbally answering simple questions
Teacher: So they outsmarted the wolf!		
Teacher: All right, Student C, will you read aloud for us.		
Student C: reads the text		Reading text aloud
Students A and C: listen to Student C read.		

Note. During a small-group lesson that lasted approximately five minutes, the teacher and her students (A, B, and C) are getting ready to read a parody of the *Three Little Pigs*. The teacher begins by describing what a parody is and provides background knowledge to help the students understand the text. The purpose or instructional focus of this lesson was to “remember and understand.” The codes used are provided in Tables S1. & S2. Numbers refer to the specific type of teacher and student talk that were coded.

Table C.2

Comparison of Teacher Measurement Models

Model	χ^2	*df*	RMSEA	CFI	*p*-value
1D	97.96	--	0.01	0.97	--
2D	91.43	2	0.01	0.98	0.04
3D	88.93	3	0.01	0.98	0.48

Table C.3

Comparison of Student Measurement Models

Model	Log-likelihood	LRT *p*-value	AIC	BIC	χ^2	*df*	RMSEA	CFI
1D	-10303	--	20660	20818	25	27	<.01	>.99
2D	-10287	$p<.001$	20630	20794	23	26	<.01	>.99
Bifac	-10273	$p<.001$	20619	20829	14	18	<.01	>.99
3D	-10286	--	20632	20808	21	24	<.01	>.99

Note. The bi-factor model evidenced the best comparative fit to the data. χ^2, RMSEA, CFI, TLI are based on the weighted least square estimator adjusted for means and variances; log-likelihood, AIC, and BIC are based on a categorical full-information maximum likelihood estimator (FIML) using the logit link. Information criteria are based on 27, 28, and 36 parameters for the one-, two-dimension, and bi-factor models, respectively. Note that all models provided excellent fit. We present the results for the general factor from the bi-factor model (see Fig 1).